

TITLE: TRANSCRIPTIONAL ADAPTOR PROTEIN

INVENTOR: William B. Solomon, Shaji  
Abraham

DOCKET NO.: 011.00250

I:\011\00250\B2

# TRANSCRIPTIONAL ADAPTOR PROTEIN

## FIELD OF THE INVENTION

The present invention relates generally to a transcriptional adaptor protein, the gene encoding the protein and uses thereof.

## BRIEF DESCRIPTION OF THE DRAWINGS

These and other features and advantages of this invention will be evident from the following detailed description of preferred embodiments when read in conjunction with the accompanying drawings in which:

Figures 1A and B show a Northern Blot Analyses of Total Cellular RNA obtained from uninduced, hemin, or TPA induced K562 cells. Figure 1A: Probes SA25, SA 21, SA11 were obtained as a partial cDNA fragments from the K562 TPA-Hemin subtracted cDNA library. SA25 hybridized to a novel 3.7 kb transcript TIG-1. SA21 hybridized to the TIMP-1 mRNA. SA11 hybridized to an EST of unknown function. Figure 1B: Fold induction of TIG-1, TIMP-1, and R35565 mRNAs with mRNAs obtained from hemin induced K562 cells taken as 1. Amounts of mRNA expression were normalized to G3PDH mRNA expression under each inducing condition.

Figures 2A and B show a Northern Blot Analysis of TPA induced K562 mRNA demonstrating time course of TIG-1 mRNA expression. Figure 2A: Representative northern blot using a partial cDNA fragment of TIG-1 as probe. Figure 2B: Bar graph of fold induction of TIG-1 mRNA at various time points after TPA induction of K562 cells. Fold induction was normalized to expression of G3PDH mRNA at each time point.

Figures 3A and B show a Northern Blot Analysis of the expression of TIG-1 mRNA in various tissues: Figure 3A: Representative Northern Blots of Tissues. Each lane

contains 2  $\mu$ g of poly A<sup>+</sup> mRNA. Figure 3B: Bar graph showing relative expression of TIG-1 mRNA in each tissue with 1 taken as level of expression in the kidney. Each bar represents the average of two independent 5 determinations of TIG-1 expression. TIG-1 mRNA expression was normalized to  $\beta$ -actin mRNA expression in each tissue.

Subal > Figure 4 illustrates the sequence of the TIG-1 cDNA: The deduced open reading frame of the TIG-1 cDNA is 10 shown. Underlined is the putative bipartite nuclear localization signal (RRMINKIDKNEDRKK). The three circles underneath TAR represent a putative protein kinase C phosphorylation site. The sequence GSSQAE is a putative casein kinase II phosphorylation site. The filled in 15 bars below the sequences NVSS and NFSV represent potential N-glycosylation sites.

Suba2 > Figure 5 shows a representation of the amino acid sequence of TIG-1 protein: The glutamine and serine/proline rich domains are noted. The glutamine 20 rich repeat sequences are shown. The putative nuclear localization signal is shown in black.

Figure 6A shows a SDS-PAGE of in vitro translation products obtained from TIG-1 transcripts. Lane 4 contains translation products of TIG-1 sense transcripts. 25 Lane 5 contains translation products of TIG-1 antisense transcripts. Lanes 2 and 3 contains translation products of NF-E2 sense and antisense transcripts respectively. Figure 6B shows a Western Blot analysis of K562 cytoplasmic and nuclear protein extracts. Each lane 30 contains 50  $\mu$ g of protein. Lanes 1 and 2 were blotted with preimmune serum which had been affinity purified with the carboxy terminal amino acids of TIG-1 protein. Lanes 3 and 4 were blotted with the supernatant fraction

of the affinity purified preimmune serum. Lanes 5 and 6 were blotted with post-immune serum affinity purified with the carboxy terminal amino acids of TIG-1 protein.

Figures 7A-C <sup>are</sup> is a Transient Transfection analysis of effect GAL4:TIG-1 fusion protein on CAT reporter gene expression. Figure 7A shows a representation of the structure of the GAL4:TIG-1 fusion proteins and CAT gene reporter plasmid construct. Figure 7B shows a representative CAT assay following 48 hours of transient expression of the GAL4:TIG-1 fusion proteins. Figure 7C shows a bar graph showing mean and standard deviation of fold induction of CAT activity in each cell lysate. The fold induction is relative to the CAT activity of uninduced K562 cells. Differences in transfection efficiency were corrected for by cotransfection with pCMVβgal expression plasmid.

#### DETAILED DESCRIPTION OF THE INVENTION

Throughout this application various publications are referenced, many in parenthesis. Full citations for each of these publications are provided at the end of the Detailed Description and throughout the Detailed Description. The disclosures of each of these publications in their entireties are hereby incorporated by reference in this application.

The term "nucleic acid", as used herein, refers to either DNA or RNA. "Nucleic acid sequence" or "polynucleotide sequence" refers to a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes both self-replicating plasmids, infectious polymers of DNA or RNA, and nonfunctional DNA or RNA.

"Isolated" nucleic acid refers to nucleic acid which has been separated from an organism in a substantially purified form (i.e. substantially free of other substances originating from that organism), and to  
5 synthetic nucleic acid.

By a nucleic acid sequence "homologous to" or "complementary to", it is meant a nucleic acid that selectively hybridizes, duplexes or binds to DNA sequences encoding the protein or portions thereof when  
10 the DNA sequences encoding the protein are present in a genomic or cDNA library. A DNA sequence which is similar or complementary to a target sequence can include sequences which are shorter or longer than the target sequence so long as they meet the functional test set  
15 forth.

Typically, the hybridization is done in a Southern blot protocol using a 0.2X SSC, 0.1% SDS, 65°C wash. The term "SSC" refers to a citrate-saline solution of 0.15M sodium chloride and 20 mM sodium citrate. Solutions are  
20 often expressed as multiples or fractions of this concentration. For example, 6X SSC refers to a solution having a sodium chloride and sodium citrate concentration of 6 times this amount or 0.9 M sodium chloride and 120 mM sodium citrate. 0.2X SSC refers to a solution 0.2  
25 times the SSC concentration or 0.03M sodium chloride and 4 mM sodium citrate.

The phrase "nucleic acid molecule encoding" refers to a nucleic acid molecule which directs the expression of a specific protein or peptide. The nucleic acid  
30 sequences include both the DNA strand sequence that is transcribed into RNA and the RNA sequence that is translated into protein or peptide. The nucleic acid molecule includes both the full length nucleic acid

sequences as well as non-full length sequences derived from the full length protein. It being further understood that the sequence includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell.

The term "located upstream" as used herein refers to linkage of a promoter upstream from a nucleic acid (DNA) sequence such that the promoter mediates transcription of the nucleic acid (DNA) sequence.

The term "vector", refers to viral expression systems, autonomous self-replicating circular DNA (plasmids), and includes both expression and nonexpression plasmids. Where a recombinant microorganism or cell is described as hosting an "expression vector," this includes both extrachromosomal circular DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by a host cell, the vector may either be stably replicated by the cells during mitosis as an autonomous structure, or the vector may be incorporated within the host's genome.

The term "plasmid" refers to an autonomous circular DNA molecule capable of replication in a cell, and includes both the expression and nonexpression types. Where a recombinant microorganism or cell is described as hosting an "expression plasmid", this includes latent viral DNA integrated into the host chromosome(s). Where a plasmid is being maintained by a host cell, the plasmid is either being stably replicated by the cell during mitosis as an autonomous structure, or the plasmid is incorporated within the host's genome.

The phrase "heterologous protein" or "recombinantly produced heterologous protein" refers to a peptide or protein of interest produced using cells that do not have an endogenous copy of DNA able to express the peptide or protein of interest. The cells produce the peptide or protein because they have been genetically altered by the introduction of the appropriate nucleic acid sequences. The recombinant peptide or protein will not be found in association with peptides or proteins and other subcellular components normally associated with the cells producing the peptide or protein.

The following terms are used to describe the sequence relationships between two or more nucleic acid molecules or polynucleotides, or between two or more amino acid sequences of peptides or proteins: "reference sequence", "comparison window", "sequence identity", "sequence homology", "percentage of sequence identity", "percentage of sequence homology", "substantial identity", and "substantial homology". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing or may comprise a complete cDNA or gene sequence.

Optimal alignment of sequences for aligning a comparison window may be conducted, for example, by the local homology algorithm of Smith and Waterman (1981), by the homology alignment algorithm of Needleman and Wunsch (1970), by the search for similarity method of Pearson and Lipman (1988), or by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.).

As applied to nucleic acid molecules or polynucleotides, the terms "substantial identity" or "substantial sequence identity" mean that two nucleic acid sequences, when optimally aligned (see above), share  
5 at least 90 percent sequence identity, preferably at least 95 percent sequence identity, more preferably at least 96, 97, 98 or 99 percent sequence identity.

"Percentage nucleotide (or nucleic acid) identity" or "percentage nucleotide (or nucleic acid) sequence  
10 identity" refers to a comparison of the nucleotides of two nucleic acid molecules which, when optimally aligned, have approximately the designated percentage of the same nucleotides. For example, "95% nucleotide identity" refers to a comparison of the nucleotides of two nucleic  
15 acid molecules which when optimally aligned have 95% nucleotide identity. Preferably, nucleotide positions which are not identical differ by redundant nucleotide substitutions (the nucleotide substitution does not change the amino acid encoded by the particular codon).

20 As further applied to nucleic acid molecules or polynucleotides, the terms "substantial homology" or "substantial sequence homology" mean that two nucleic acid sequences, when optimally aligned (see above), share at least 90 percent sequence homology, preferably at  
25 least 95 percent sequence homology, more preferably at least 96, 97, 98 or 99 percent sequence homology.

"Percentage nucleotide (or nucleic acid) homology" or "percentage nucleotide (or nucleic acid) sequence  
30 homology" refers to a comparison of the nucleotides of two nucleic acid molecules which, when optimally aligned, have approximately the designated percentage of the same nucleotides or nucleotides which are not identical but differ by redundant nucleotide substitutions (the



As further applied to polypeptides, the terms "substantial homology" or "substantial sequence homology" mean that two peptide sequences, when optimally aligned, 30 such as by the programs GAP or BESTFIT using default gap, share at least 90 percent sequence homology, preferably at least 95 percent sequence homology, more preferably at least 96, 97, 98 or 99 percent sequence homology.

"Percentage amino acid homology" or "percentage amino acid sequence homology" refers to a comparison of the amino acids of two polypeptides which, when optimally aligned, have approximately the designated percentage of the same amino acids or conservatively substituted amino acids. For example, "95% amino acid homology" refers to a comparison of the amino acids of two polypeptides which when optimally aligned have 95% amino acid homology. As used herein, homology refers to identical amino acids or residue positions which are not identical but differ only by conservative amino acid substitutions. For example, the substitution of amino acids having similar chemical properties such as charge or polarity are not likely to affect the properties of a protein. Examples include glutamine for asparagine or glutamic acid for aspartic acid.

The phrase "substantially purified" or "isolated" when referring to a protein (or peptide), means a chemical composition which is essentially free of other cellular components. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein (or peptide) which is the predominant species present in a preparation is substantially purified. Generally, a substantially purified or isolated protein (or peptide) will comprise more than 80% of all macromolecular species present in the preparation. Preferably, the protein (or peptide) is purified to represent greater than 90% of all macromolecular species present. More preferably the protein (or peptide) is purified to greater than 95%, and

most preferably the protein (or peptide) is purified to essential homogeneity, wherein other macromolecular species are not detected by conventional techniques. As used herein, a "substantially purified" or "isolated" protein (or peptide) can be synthetically or chemically produced, or recombinantly produced. A "substantially purified" or "isolated" protein or peptide as used herein is not intended to include a protein or peptide separated from an organism.

10 "Biological sample" or "sample" as used herein refers to any sample obtained from a living organism or from an organism that has died. Examples of biological samples include body fluids and tissue specimens.

High stringent hybridization conditions are selected at about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is at least about 0.02 molar at pH 7 and the temperature is at least about 60°C. As other factors may significantly affect the stringency of hybridization, including, among others, base composition and size of the complementary strands, the presence of organic solvents, i.e. salt or formamide concentration, and the extent of base mismatching, the combination of parameters is more important than the absolute measure of any one. High stringency may be attained, for example, by overnight hybridization at about 68°C in a 6X SSC solution, washing at room temperature with 6X SSC solution, followed by washing at about 68°C in a 6X SSC solution then in a 0.6X SSX solution.

Hybridization with moderate stringency may be attained, for example, by: 1) filter pre-hybridizing and hybridizing with a solution of 3X sodium chloride, sodium citrate (SSC), 50% formamide, 0.1M Tris buffer at pH 7.5, 5X Denhardt's solution; 2) pre-hybridization at 37°C for 4 hours; 3) hybridization at 37°C with amount of labeled probe equal to 3,000,000 cpm total for 16 hours; 4) wash in 2X SSC and 0.1% SDS solution; 5) wash 4X for 1 minute each at room temperature and 4X at 60°C for 30 minutes each; and 6) dry and expose to film.

The phrase "selectively hybridizing to" refers to a nucleic acid molecule that hybridizes, duplexes or binds only to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. By selectively hybridizing it is meant that a nucleic acid molecule binds to a given target in a manner that is detectable in a different manner from non-target sequence under moderate, or more preferably under high, stringency conditions of hybridization. "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid molecule. Proper annealing conditions depend, for example, upon a nucleic acid molecule's length, base composition, and the number of mismatches and their position on the molecule, and must often be determined empirically. For discussions of nucleic acid molecule (probe) design and annealing conditions, see, for example, Sambrook et al. 1989.

It will be readily understood by those skilled in the art and it is intended here, that when reference is made to particular sequence listings, such reference includes sequences which substantially correspond to its complementary sequence and those described including

allowances for minor sequencing errors, single base changes, deletions, substitutions and the like, such that any such sequence variation corresponds to the nucleic acid sequence of the peptide/protein to which the  
5 relevant sequence listing relates.

The DNA molecules of the subject invention also include DNA molecules coding for protein analogs, fragments or derivatives of the protein which differ from naturally-occurring forms (the naturally-occurring  
10 protein) in terms of the identity or location of one or more amino acid residues (deletion analogs containing less than all of the residues specified for the protein, substitution analogs wherein one or more residues specified are replaced by other residues, and addition  
15 analogs wherein one or more amino acid residues are added to a terminal or medial portion of the protein) and which share the function of the naturally-occurring form. These molecules include: the incorporation of codons "preferred" for expression by selected non-mammalian  
20 hosts; the provision of sites for cleavage by restriction endonuclease enzymes; and the provision of additional initial, terminal or intermediate DNA sequences that facilitate construction of readily expressed vectors.

As used herein, a "peptide" refers to an amino acid  
25 sequence of three to one hundred amino acids, and therefore an isolated peptide that comprises an amino acid sequence is not intended to cover amino acid sequences of greater than 100 amino acids. Preferably, the peptides that can be identified and used in  
30 accordance with the subject invention (whether they be mimotope or anti-mimotope peptides) are less than 50 amino acids in length, and more preferably the peptides

are five to 20 amino acids in length or 20-40 amino acids in length.

The peptides can contain any naturally-occurring or non-naturally-occurring amino acids, including the D-form of the amino acids, amino acid derivatives and amino acid mimics, so long as the desired function and activity of the peptide is maintained. The choice of including an (L)- or a (D)-amino acid in the peptides depends, in part, on the desired characteristics of the peptide. For example, the incorporation of one or more (D)-amino acids can confer increased stability on the peptide and can allow a peptide to remain active in the body for an extended period of time. The incorporation of one or more (D)-amino acids can also increase or decrease the pharmacological activity of the peptide.

The peptides may also be cyclized, since cyclization may provide the peptides with superior properties over their linear counterparts.

As used herein, the terms "amino acid mimic" and "mimetic" mean an amino acid analog or non-amino acid moiety that has the same or similar functional characteristic of a given amino acid. For instance, an amino acid mimic of a hydrophobic amino acid is one which is non-polar and retains hydrophobicity, generally by way of containing an aliphatic chemical group. By way of further example, an arginine mimic can be an analog of arginine which contains a side chain having a positive charge at physiological pH, as is characteristic of the guanidinium side chain reactive group of arginine.

In addition, modifications to the peptide backbone and peptide bonds thereof are also encompassed within the scope of amino acid mimic or mimetic. Such modifications can be made to the amino acid, derivative thereof,

non-amino acid moiety or the peptide either before or after the amino acid, derivative thereof or non-amino acid moiety is incorporated into the peptide. What is critical is that such modifications mimic the peptide backbone and bonds which make up the same and have substantially the same spacial arrangement and distance as is typical for traditional peptide bonds and backbones. An example of one such modification is the reduction of the carbonyl(s) of the amide peptide backbone to an amine. A number of reagents are available and well known for the reduction of amides to amines such as those disclosed in Wann et al., JOC, 46:257 (1981) and Raucher et al., Tetrahedron. Lett., 21:14061 (1980). An amino acid mimic is, therefor, an organic molecule that retains the similar amino acid pharmacophore groups as is present in the corresponding amino acid and which exhibits substantially the same spatial arrangement between functional groups.

The substitution of amino acids by non-naturally occurring amino acids and amino acid mimics as described above can enhance the overall activity or properties of an individual peptide based on the modifications to the backbone or side chain functionalities. For example, these types of alterations to the amino acid substituents and peptides can enhance the peptide's stability to enzymatic breakdown and increase biological activity. Modifications to the peptide backbone similarly can add stability and enhance activity.

One skilled in the art, using the above sequences or formulae, can easily synthesize the peptides. Standard procedures for preparing synthetic peptides are well known in the art. The novel peptides can be synthesized using: the solid phase peptide synthesis (SPPS) method of

000000 " 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

Merrifield (J. Am. Chem. Soc., 85:2149 (1964)) or modifications of SPPS; or, the peptides can be synthesized using standard solution methods well known in the art (see, for example, Bodanzsky, M., Principles of Peptide Synthesis, 2nd revised ed., Springer-Verlag (1988 and 1993)). Alternatively, simultaneous multiple peptide synthesis (SMPS) techniques well known in the art can be used. Peptides prepared by the method of Merrifield can be synthesized using an automated peptide synthesizer such as the Applied Biosystems 431A-01 Peptide Synthesizer (Mountain View, Calif.) or using the manual peptide synthesis technique described by Houghten, Proc. Natl. Acad. Sci., USA 82:5131 (1985).

With these definitions in mind, the subject invention provides isolated nucleic acid molecules and proteins encoded by the isolated nucleic acid molecules. The proteins have transcriptional activation activity, i.e. they activate transcription. The nucleic acid molecules can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA, including messenger RNA or mRNA), genomic or recombinant, biologically isolated or synthetic.

The DNA molecules can be a cDNA molecule, which is a DNA copy of a messenger RNA (mRNA) encoding the protein.

25 *Sub 2302* An example of the protein is the protein encoded by the nucleotide sequence as shown in SEQ ID NO:1 (this is the open reading frame). The amino acid sequence encoded by this nucleotide sequence is shown in SEQ ID NO:3. The full nucleotide sequence is as shown in SEQ ID NO:2. 30 The amino acid sequence encoded by this nucleotide sequence is shown in SEQ ID NO:4.

The invention also provides an oligonucleotide that is complementary to at least a portion of the mRNA



encoding the protein. Oligonucleotides can be RNA or single-stranded DNA, and can be complementary to the entire mRNA molecule encoding the protein (i.e. of the same nucleotide length as the entire molecule). It may be desirable, however, to work with a shorter molecule. In this instance, the oligonucleotide can be complementary to a portion of the entire mRNA molecule encoding the protein. These shorter oligonucleotides are capable of hybridizing to the mRNA encoding the entire molecule, and preferably consist of about twenty to about one hundred nucleotides. These oligonucleotides can be used to reduce levels of proteins having transcriptional activating activity, by introducing into cells an RNA or single-stranded DNA molecule that is complementary to at least a portion of the mRNA of the protein (i.e. by introducing the oligonucleotide). The oligonucleotide can base-pair with the mRNA of the protein, preventing translation of the mRNA into protein. Thus, an oligonucleotide can prevent translation of mRNA encoding the protein into a functional protein. It may be desirable to place the oligonucleotide downstream and under the control of a particular promoter, so that the oligonucleotide will prevent translation of mRNA encoding the protein only in cells in which the particular promoter functions.

More particularly, an oligonucleotide complementary to at least a portion of mRNA encoding a transcriptional activator protein can be used to decrease expression of a functional channel. A cell with a first level of expression of a functional transcriptional activator protein is selected, and then the oligonucleotide is introduced into the cell. The oligonucleotide blocks expression of functional transcriptional activator

protein, resulting in a second level of expression of a functional transcriptional activator protein in the cell. The second level is less than the initial first level.

Oligonucleotides can be introduced into cells by any  
5 suitable means. In one embodiment, the oligonucleotide RNA molecule is injected directly into the cellular cytoplasm, where the RNA interferes with translation. A vector may also be used for introduction of the oligonucleotide into a cell. Such vectors include  
10 various plasmid and viral vectors. For a general discussion of oligonucleotides such as antisense molecules and their use, see Han et al. 1991 and Rossi 1995.

The nucleic acid molecules of the subject invention  
15 can be expressed in suitable host cells using conventional techniques. Any suitable host and/or vector system can be used to express the transcriptional activator protein.

Techniques for introducing the nucleic acid  
20 molecules into the host cells may involve the use of expression vectors which comprise the nucleic acid molecules. These expression vectors (such as plasmids and viruses; viruses including bacteriophage) can then be used to introduce the nucleic acid molecules into  
25 suitable host cells. For example, DNA encoding the transcriptional activator protein can be injected into the nucleus of a host cell or transformed into the host cell using a suitable vector, or mRNA encoding the transcriptional activator protein can be injected  
30 directly into the host cell, in order to obtain expression of the transcriptional activator protein in the host cell.

Various methods are known in the art for introducing nucleic acid molecules into host cells. One method is microinjection, in which DNA is injected directly into the nucleus of cells through fine glass needles (or RNA is injected directly into the cytoplasm of cells). Alternatively, DNA can be incubated with an inert carbohydrate polymer (dextran) to which a positively charged chemical group (DEAE, for diethylaminoethyl) has been coupled. The DNA sticks to the DEAE-dextran via its negatively charged phosphate groups. These large DNA-containing particles stick in turn to the surfaces of cells, which are thought to take them in by a process known as endocytosis. Some of the DNA evades destruction in the cytoplasm of the cell and escapes to the nucleus, where it can be transcribed into RNA like any other gene in the cell. In another method, cells efficiently take in DNA in the form of a precipitate with calcium phosphate. In electroporation, cells are placed in a solution containing DNA and subjected to a brief electrical pulse that causes holes to open transiently in their membranes. DNA enters through the holes directly into the cytoplasm, bypassing the endocytotic vesicles through which they pass in the DEAE-dextran and calcium phosphate procedures. DNA can also be incorporated into artificial lipid vesicles, liposomes, which fuse with the cell membrane, delivering their contents directly into the cytoplasm. In an even more direct approach, DNA is absorbed to the surface of tungsten microprojectiles and fired into cells with a device resembling a shotgun.

Several of these methods, microinjection, electroporation, and liposome fusion, have been adapted to introduce proteins into cells. For review, see

Further methods for introducing nucleic acid molecules into cells involve the use of viral vectors.

As indicated, some of these methods of transforming a cell require the use of an intermediate plasmid vector. U.S. Patent No. 4,237,224 to Cohen and Boyer describes the production of expression systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase. These recombinant plasmids are then introduced by means of transformation and replicated in unicellular cultures including procaryotic organisms and eucaryotic cells grown in tissue culture. The DNA sequences are cloned into the plasmid vector using standard cloning procedures known in the art, as described by Sambrook et al. (1989).

Having identified the nucleic acid molecules encoding transcriptional activator proteins and methods for expressing the transcriptional activator proteins encoded thereby, the invention further provides methods of screening a substance (for example, a compound or inhibitor) for the ability of the substance to modify transcriptional activator protein function. In one embodiment, the method comprises introducing a nucleic

acid molecule encoding the transcriptional activator protein into a host cell, and expressing the transcriptional activator protein encoded by the molecule in the host cell. The cell is then exposed to a  
5 substance and evaluated to determine if the substance modifies the function of the transcriptional activator protein. In another embodiment, an isolated transcriptional activator protein is exposed to the substance for evaluation of whether the substance  
10 modifies the function of the transcriptional activator protein. From these evaluations, substances effective in altering the function of the transcriptional activator protein can be found. Such agents may be agonists or antagonists, with antagonists being preferred herein.  
15 The evaluation of a cell to determine if the substance modifies the function of the transcriptional activator protein can be by any means known in the art. The evaluation can comprise the direct monitoring of expression of transcriptional activator protein in the  
20 host cell, or the evaluation can be indirect.

The nucleic acid molecules of the subject invention can be used either as probes or for the design of primers to obtain DNA encoding other transcriptional activator proteins by either cloning and colony/plaque  
25 hybridization or amplification using the polymerase chain reaction (PCR).

Specific probes derived from SEQ ID NO:1 can be employed to identify colonies or plaques containing cloned DNA encoding a member of the transcriptional  
30 activator protein family using known methods (see Sambrook et al. 1989). One skilled in the art will recognize that by employing such probes under high stringency conditions (for example, hybridization at 42°C

with 5X SSPC and 50% formamide, washing at 50-65°C with 0.5X SSPC), sequences having regions which are greater than 90% homologous or identical to the probe can be obtained. Sequences with lower percent homology or  
5 identity to the probe, which also encode transcriptional activator proteins, can be obtained by lowering the stringency of hybridization and washing (e.g., by reducing the hybridization and wash temperatures or reducing the amount of formamide employed).

10 More particularly, in one embodiment, the method comprises selection of a DNA molecule encoding a transcriptional activator protein, or a fragment thereof, the DNA molecule having a nucleotide sequence as shown in SEQ ID NO:1, and designing an oligonucleotide probe for  
15 transcriptional activator protein based on the nucleotide sequence of the selected DNA molecule. A genomic or cDNA library of an organism is then probed with the oligonucleotide probe, and clones are obtained from the library that are recognized by the oligonucleotide probe  
20 so as to obtain DNA encoding another transcriptional activator protein.

Specific primers derived from SEQ ID NO:1 can be used in PCR to amplify a DNA sequence encoding a member of the transcriptional activator protein family using  
25 known methods (see Innis et al. 1990). One skilled in the art will recognize that by employing such primers under high stringency conditions (for example, annealing at 50-60°C, depending on the length and specific nucleotide content of the primers employed), sequences  
30 having regions greater than 75% homologous or identical to the primers will be amplified.

More particularly, in a further embodiment the method comprises selection of a DNA molecule encoding

transcriptional activator protein, or a fragment thereof,  
the DNA molecule having a nucleotide sequence as shown in  
SEQ ID NO:1, designing degenerate oligonucleotide primers  
based on the nucleotide sequence of the selected DNA  
5 molecule, and employing such primers in the polymerase  
chain reaction using as a template a DNA sample to be  
screened for the presence of transcriptional activator  
protein-encoding sequences. The resulting PCR products  
can be isolated and sequenced to identify DNA fragments  
10 that encode polypeptide sequences corresponding to the  
targeted region of transcriptional activator protein.

Various modifications of the nucleic acid and amino  
acid sequences disclosed herein are covered by the  
subject invention. These varied sequences still encode a  
15 functional transcriptional activator protein. The  
invention thus further provides an isolated nucleic acid  
molecule encoding a transcriptional activator protein,  
the nucleic acid molecule encoding a first amino acid  
sequence having at least 90% amino acid identity to a  
20 second amino acid sequence, the second amino acid  
sequence as shown in SEQ ID NO:3. In further  
embodiments, the first amino acid sequence has at least  
95%, 96%, 97%, 98%, or 99% amino acid identity to SEQ ID  
NO:3.

25 The invention further provides an isolated DNA  
oligomer capable of hybridizing to the nucleic acid  
molecule encoding the transcriptional activator protein  
according to the subject invention. Such oligomers can  
be used as probes in a method of detecting the presence  
30 of transcriptional activator protein in a sample. More  
particularly, a sample can be contacted with the DNA  
oligomer and the DNA oligomer will hybridize to any  
transcriptional activator protein present in the sample,

forming a complex therewith. The complex can then be detected, thereby detecting presence of transcriptional activator protein in the sample.

The complex can be detected using methods known in the art. Preferably, the DNA oligomer is labeled with a detectable marker so that detection of the marker after the DNA oligomer hybridizes to any transcriptional activator protein in the sample (wherein non-hybridized DNA oligomer has been washed away) is detection of the complex. Detection of the complex indicates the presence of transcriptional activator protein in the sample. As will be readily apparent to those skilled in the art, such a method could also be used quantitatively to assess the amount of transcriptional activator protein in a sample.

For detection, the oligomers can be labeled with, for example, a radioactive isotope, biotin, an element opaque to X-rays, or a paramagnetic ion. Radioactive isotopes are commonly used and are well known to those skilled in the art. Representative examples include indium-111, technetium-99m, and iodine-123. Biotin is a standard label which would allow detection of the biotin labeled oligomer with avidin. Paramagnetic ions are also commonly used and include, for example, chelated metal ions of chromium (III), manganese (II), and iron (III). When using such labels, the labeled DNA oligomer can be imaged using methods known to those skilled in the art. Such imaging methods include, but are not limited to, X-ray, CAT scan, PET scan, NMRI, and fluoroscopy. Other suitable labels include enzymatic labels (horseradish peroxidase, alkaline phosphatase, etc.) and fluorescent labels (such as FITC or rhodamine, etc.).



006644 092200

The invention further provides an isolated transcriptional activator protein. The protein is preferably encoded by a nucleotide sequence as shown in SEQ ID NO:1. The protein preferably has an amino acid sequence as shown in SEQ ID NO:3. Further provided is an isolated transcriptional activator protein encoded by a first amino acid sequence having at least 90% amino acid identity to a second amino acid sequence, the second amino acid sequence as shown in SEQ ID NO:3. In further embodiments, the first amino acid sequence has at least 95%, 96%, 97%, 98%, or 99% amino acid identity to SEQ ID NO:3.

The invention further provides an antibody or fragment thereof specific for the transcriptional activator protein of the subject invention. Antibodies of the subject invention include polyclonal antibodies and monoclonal antibodies capable of binding to the transcriptional activator protein, as well as fragments of these antibodies, and humanized forms. Humanized forms of the antibodies of the subject invention may be generated using one of the procedures known in the art such as chimerization. Fragments of the antibodies of the present invention include, but are not limited to, the Fab, the F(ab')<sub>2</sub>, and the Fc fragments.

The invention also provides hybridomas which are capable of producing the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (see Campbell 1984 and St. Groth et al. 1980). Any animal (mouse, rabbit, etc.) which is known to produce

10 The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to, coupling  
15 the antigen with a heterologous protein (such as a globulin or beta-galactosidase) or through the inclusion of an adjuvant during immunization.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These  
25 include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz et al. 1988).

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is

screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

In accordance with the above discussion, the subject invention further provides a method of producing an antibody specific for a transcriptional activator protein in a host. The method comprises selecting the isolated transcriptional activator protein or an antigenic portion thereof and introducing the selected transcriptional activator protein or antigenic portion thereof into a host to induce production of an antibody specific for transcriptional activator protein in the host.

The present invention further provides the above-described antibodies in detectably labeled form. Antibodies can be detectably labeled through the use of radioisotopes, affinity labels (such as biotin, avidin, etc.), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, etc.), fluorescent labels (such as FITC or rhodamine, etc.), paramagnetic atoms, etc. Procedures for accomplishing such labeling are well known in the art, for example see Sternberger et al. 1970, Bayer et al. 1979, Engval et al. 1972, and Goding 1976.

The labeled antibodies or fragments thereof of the present invention can be used for in vitro, in vivo, and in situ assays to identify cells or tissues which express transcriptional activator protein, to identify samples containing transcriptional activator protein, or to detect the presence of transcriptional activator protein in a sample. More particularly, the antibodies or fragments thereof can thus be used to detect the presence of transcriptional activator protein in a sample, by contacting the sample with the antibody or fragment thereof. The antibody or fragment thereof binds to any transcriptional activator protein present in the sample,

forming a complex therewith. The complex can then be detected, thereby detecting the presence of transcriptional activator protein in the sample. As will be readily apparent to those skilled in the art, such a method could also be used quantitatively to assess the amount of transcriptional activator protein in a sample. As should also be readily apparent, such an antibody may also be used to decrease levels of functional transcriptional activator protein, by blocking the protein.

#### EXAMPLE

A subtractive hybridization protocol was used to identify novel expressed sequence tags (ESTs) corresponding to mRNAs whose expression was induced upon exposure of the human leukemia cell line K562 to the phorbol ester 12-O-tetradecanolyphorbol-13-acetate (TPA).

The complete open reading frame of one of the novel ESTs, named *TIG-1*, was obtained by screening K562 cell and placental cDNA libraries. The deduced open reading frame of the *TIG-1* cDNA encodes for a glutamine repeat rich protein with a predicted molecular weight of 63 kDa. The predicted open reading frame also contains a consensus bipartite nuclear localization signal, though no specific DNA binding domain was found.

The corresponding *TIG-1* mRNA is ubiquitously expressed. Placental tissue expresses the *TIG-1* mRNA 200 fold times more than the lowest expressing tissues such as kidney and lung. There is also preferential *TIG-1* mRNA expression in cells of bone marrow lineage.

*In vitro* transcription/translation of the *TIG-1* cDNA yielded a polypeptide with an apparent molecular

weight of 97 kDa. Using polyclonal antibodies obtained from a rabbit immunized with the carboxy terminal portion of bacterially expressed TIG-1 protein, a polypeptide with molecular weight of 97 kDa was identified by Western blot analyses of protein lysates obtained from K562 cells.

Cotransfection assays of K562 cells, using a *GAL4-TIG-1* fusion gene and *GAL4* operator-CAT, indicate that the TIG-1 protein may have transcriptional regulatory activity when tethered to DNA. It was hypothesized that this novel glutamine rich protein participates in a protein complex that regulates gene transcription. It has been demonstrated by Naar et. al., 1999 that the amino acid sequences of peptide fragments obtained from a polypeptide found in a complex of proteins that alters chromatin structure (ARC) are identical to portions of the deduced open reading frame of TIG-1 mRNA.

## 20 ~~Introduction~~

The K562 human leukemia cell line can be made to differentiate to a variety of myeloid derived lineages. Induction with hemin leads to expression of markers of erythroid differentiation, while exposure to the phorbol ester 12-O-tetradecanolyphorbol-13-acetate (TPA) leads to extinction of the erythroid phenotype and induction of a megakaryocytic phenotype. Previous work in many laboratories, has identified a set of erythroid specific genes, such as the globins and glycophorin A, whose expression was increased upon hemin induction of K562 cells (Andersson et. al., 1979, Charnay and Maniatis, 1983, Dean et. al., 1983). Gene expression during TPA induced megakaryocytic differentiation of K562 cells has

been less well studied (Alitalo et. al. 1988, Alitalo et. al. 1990, Lin et al. 1994, Lumelsky and Forget, 1991).

Since the K562 cell line can serve as a model for pluripotent bone marrow stem cell differentiation; it is of interest to identify novel genes preferentially expressed in myeloid lineage cells. Subtractive hybridization protocol, first introduced by Wang and Brown, was used to identify cDNAs corresponding to novel mRNAs whose expression was increased after K562 cell induction with TPA.

The molecular cloning, expression, and functional characterization of *TIG-1*, a novel human CAG repeat rich cDNA, whose corresponding mRNA contains a deduced open reading frame encoding for a glutamine repeat rich protein with a consensus bipartite nuclear localization signal is reported. The 3.7 kb mRNA encoding for this protein, though constitutively expressed, is highly preferentially expressed in placental and bone marrow lineage cells. Using rabbit polyclonal antibodies raised to the bacterially expressed carboxy terminal portion of the deduced open reading frame of *TIG-1*, Western blot analyses of K562 cell protein extracts show that *TIG-1* is found in the cytoplasm and nucleus with an apparent molecular weight of 97 kDa.

In transient transfection assays of K562 cells transfected with a *GAL4:TIG-1* fusion gene in an eukaryotic expression vector, evidence is presented that the protein encoded by this novel cDNA has transcriptional regulatory activity when tethered to DNA by the Gal 4 DNA binding domain.

## Material and Methods

### Cell Culture:

K562 cells were grown in RPMI 1640 medium supplemented with 10% heat inactivated fetal bovine serum and incubated at 37° C in the presence of 5% CO<sub>2</sub>. Cells were induced with 20  $\mu$ M hemin for 48 hours or 34 nM TPA for 72 hours.

*Construction of expressed sequence tag library by subtraction of hemin induced cDNAs from TPA induced cDNAs derived from mRNAs obtained from K562 cells:*

10 Poly A<sup>+</sup> mRNAs were isolated from K562 cells induced with either 20  $\mu$ M hemin for 72 hours or 34 nM TPA for 48 hours. The corresponding cDNAs were synthesized using a cDNA construction kit (Amersham, Chicago, IL).

15 *Ray* > The protocol elaborated by Wang and Brown 1991 with some modification. Long cDNAs were digested with the restriction endonuclease Rsa 1 and then ligated to linker DNAs (5'GAATTCAGATCTCCCGGGTCACCGC3' and 5'TGACCCGGGAGATCGAATTC3'). Linkered cDNA fragments were amplified by PCR. PCR amplified cDNA fragments

20 constructed from the TPA induced mRNAs were used as "tracer" cDNAs, while a five fold molar excess of biotinylated PCR amplified cDNA fragments constructed from hemin induced mRNAs were used as "driver" cDNAs to produce a EST library that was highly enriched in cDNA

25 fragments generated from the TPA induced K562 cells.

The subtractive hybridization reaction was performed as follows:

10  $\mu$ g of cDNA fragments produced from hemin induced K562 mRNAs were mixed with 10  $\mu$ g of photoprobe biotin™

30 (Vector Laboratories, Burlingame, CA ) kept on ice and was exposed to a 270 Watt sunlamp kept 10 cm above the mixture for 15 minutes. The biotinylation procedure was performed a second time. The reaction was then stopped

by neutralization with an equal volume of 0.1 M Tris HCl (pH 9.0) and the final volume was adjusted to 100  $\mu$ l with distilled water.

Unreacted biotin species were then extracted with  
5 the addition of an equal volume of 2-Butanol and the extraction was repeated twice. The aqueous phase containing the biotinylated DNA was precipitated with 1/10 volume of 1 M NaCl and 2 Volumes of ethanol. The biotinylated cDNA pool was resuspended in 10  $\mu$ l of  
10 Tris:EDTA (pH 7.6).

5  $\mu$ g of biotinylated cDNA "driver" fragments produced from hemin induced mRNAs were mixed with 1  $\mu$ g of "tracer" cDNA fragments produced from TPA induced K562 mRNAs and precipitated with ethanol; then resuspended in  
15 20  $\mu$ l of TE pH 8.0. This mixture was denatured by boiling in a water bath for 3 minutes, centrifuged briefly and then an equal volume of 2x hybridization buffer (50 mM HEPES (pH 7.5), 2 mM EDTA, 1.5 M NaCl) was added and overlaid with mineral oil. The hybridization  
20 mixture was denatured again in a boiling water bath for 3 minutes and then hybridized for 2 hours at 63° C.

The hybridization reaction was stopped with the addition of 160  $\mu$ L of HE buffer solution (10 mM HEPES, 1mM EDTA pH 7.6 prewarmed to 55° C) and the aqueous phase  
25 transferred to an eppendorf tube prewarmed at 55° C for five minutes. The biotinylated cDNAs were complexed by the addition of 5 $\mu$ L of streptavidin solution (10 mg/ml of streptavidin in 0.15 M NaCl, 10 mM HEPES pH 7.6, 1mM EDTA) and the reaction mix was incubated at room  
30 temperature for 20 minutes. The streptavidin:biotin complex was extracted with equal volumes of phenol:chloroform (1:1) prewarmed to 37° C. Phenol:chloroform extractions were repeated until no



visible DNA complex was noted at interphase. Then the aqueous phase containing the remaining subtracted cDNAs was extracted with chloroform, ethanol precipitated and resuspended in 20  $\mu$ L of TE (pH 8.0).

- 5       The entire procedure described above was repeated using the subtracted cDNAs obtained from the short hybridization reaction as "tracer", with the remaining biotinylated driver cDNA fragments; but this time the hybridization mixture was hybridized for 24 hours at 63°
- 10 C.

      The subtracted cDNAs enriched for cDNAs derived from TPA induced K562 mRNAs were then amplified by PCR for thirty cycles (94° C 1 minute, 50° C 1 minute, 72° C 3 minutes). The resultant amplified enriched cDNA pool was

15 then digested with the restriction endonuclease EcoR1, ligated into the plasmid Bluescript pBS KS II+, and transformed into XL-blue 1 cells by electroporation. Transformed cells were plated to produce the expressed sequence tag library enriched for inserts derived from

20 TPA induced K562 cell mRNAs.

#### *Expressed sequence tag library sequence analysis*

      Expressed sequence tag inserts in the plasmid pBS KS+ were flanked by the T3 and T7 promoters. Inserts

25 were directly sequenced using T3 or T7 sequencing primers. Sequence analysis of the inserts was performed on GenBank, EMBL, PIR, and Swissprot databases using FASTA and BLAST algorithms.

#### *30 Screening of cDNA libraries*

      The 193 bp EST fragment from clone pSA25 was used as a probe to screen K562 or placental cell 1 GT11 cDNA libraries. Using standard screening techniques we

identified a K562 cell clone that contained a 1.5 kb insert which contained the 193 bp sequence of the screening probe. This 1.5 kb cDNA fragment was used to screen the placental phage clones and identified 5 individual clones containing 2.1 and 2.5 kb of cDNA sequence. DNA sequences of these partial cDNAs were obtained by manual and automated sequencing.

#### *Northern Blots:*

10 Total RNA was isolated from uninduced, hemin, or TPA induced K562 cells with TRI reagent (Molecular Research Center, Inc., Cincinnati, OH) according to the manufacturer's protocol. Twenty  $\mu$ g of total RNA was size fractionated in a 1.5% formaldehyde agarose gel. The RNA 15 was transferred onto a nylon membrane overnight by capillary action. The nylon membrane was washed briefly in 6X SSC and then UV cross-linked. Membranes were stored at 4° C. Multiple organ and immune tissue blots were purchased from Clontech laboratories (Palo Alto, 20 CA).

All blots were probed with the 193 bp EST fragment of TIG-1. Expression of TIG-1 mRNA in K562 cell Northern blots was normalized to expression of G3PDH mRNA; for blots obtained from Clontech, expression of TIG-1 was 25 normalized to expression of  $\beta$ -actin mRNA.

#### *In vitro transcription and translation*

A plasmid containing 3.2 kb of the TIG-1 cDNA was linearized by restriction enzyme digestion and then 30 transcribed with either T7 or T3 RNA polymerase to generate sense or antisense transcripts. Transcripts produced *in vitro* were analyzed on a 1.5% formaldehyde agarose gel.

A rabbit reticulocyte lysate obtained from Promega (Madison, WI) was used for the *in vitro* translation reaction. *In vitro* translation products were analyzed on a 10% SDS-polyacrylamide gel that was stained with Coomassie Blue.

*Production of antibodies to the carboxy terminal portion of the TIG-1 protein*

PCR primers were used to amplify the coding region of TIG-1 protein from amino acids 394 to 579. This fragment was cloned in frame 3' to the coding sequence for glutathione S-transferase (GST). The resultant plasmid was transformed into *E. coli* which were grown and induced with 0.5 mM IPTG to overexpress the GST:TIG-1 fusion protein. This protein was affinity purified with glutathione agarose beads following the manufacturers directions. Confirmation that a GST fusion protein was obtained by performing a Western Blot using a monoclonal antibody directed to the GST portion of the fusion protein. A fusion protein of the correct expected molecular weight was detected.

The GST:TIG-1 fusion protein was used to immunize a rabbit (CoCalico Biologics, Reamstown, PA). Antiserum was obtained six weeks, 12, and 18 weeks after the first injection of the protein.

*Affinity purification of TIG-1 antiserum*

One ml of post immune serum was incubated with GST protein at 4°C overnight. The reaction was then centrifuged at 3000 rpm for 5 minutes and the supernatant was transferred to a fresh tube. This supernatant was incubated at 4°C for 12 hours with a nitrocellulose membrane which contained the 46 kDa GST:TIG-1 fusion

protein transferred to the nitrocellulose membrane following SDS-PAGE electrophoresis. The membrane was washed in PBS at room temperature for 15' repeated x 3; then affinity purified antibody was eluted from the  
5 membrane by addition of elution buffer (0.1 M glycine, 0.1 M NaCl pH 2.5) and shaking at room temperature for 20'. The eluate was transferred to a fresh tube and neutralized with 1 M Tris pH 8.8 to a pH of 7.5. Bovine serum albumin (BSA) was added to a final concentration of  
10 5 mg/ml. The affinity purified antibodies were stored at 4°C till further use.

*Western Blot analyses of K562 cytoplasmic and nuclear extracts with affinity purified antibodies:*

15 Nuclear extracts were made from uninduced, hemin, and TPA induced K562 cells using the Dignam protocol (Dignam et. al. 1983). Cytoplasmic extracts were obtained following Dounce homogenization and centrifugation to produce the nuclear pellet.  
20 Cytoplasmic and Dignam nuclear extracts were stored at -70°C till further use.

50 µg aliquots of protein from the cytoplasmic or nuclear extracts were size fractionated on a 10% SDS-PAGE gel and transferred to a nitrocellulose membrane.  
25 Membranes were incubated for 1 hour at room temperature with either a 1:100 dilution of preimmune serum or 1:5 dilutions of affinity purified pre- or post -immune serum. The membrane was washed and incubated with goat antirabbit antibody followed by a chemi-luminescent  
30 reaction used to identify protein bands that were specifically recognized by post-immune serum.

*Transient transfection experiments*

5 x 10<sup>7</sup> cells were used for each transfection. All transfections contained 5 µg of the reporter plasmid pGALx50P E1B CAT, 2 µg of pCMVβ gal (to assay for transfection efficiency), and a test plasmid in either a 1:1 or 1:5 molar ratio with the reporter plasmid. Plasmid DNAs were transiently transfected into K562 cells via electroporation at 250V and 960 µF with a Bio-Rad (Hercules, CA) Gene pulser. After resting on ice for 30 minutes, each electroporation reaction was resuspended in 10 ml of fresh RPMI 1640 medium and incubated for 12 hours at 37° C. Then each dish of transfected cells was split into three equal volumes and replated using either TPA or hemin as an inducing agent.

Forty eight hours after transient transfection whole cell freeze-thaw lysates were prepared and stored till further use at -20° C. Protein concentration of lysates were measured with the Bradford Reagent. Equal amounts of protein lysates from each dish of transfected cells was used to measure β-gal and CAT protein activities via standard techniques.

## Results and Discussion

### *Identification of expressed sequence tag fragments corresponding to TPA induced mRNAs*

Following two rounds of subtractive hybridization, the recombinant cDNA clones were isolated and their sequences searched for possible homology to genes in the NIH database. To confirm that the subtractive hybridization yielded cDNA clones corresponding to mRNAs whose expression increased following TPA induction of K562 cells, the expression of three randomly picked clones: SA11, SA21, and SA25 was analyzed by Northern

The cDNA fragment (SA21) was found to be identical to a partial sequence for the tissue inhibitor of metalloproteinases gene-1 (Alitalo et. al. 1990) which showed (in comparison to hemin-induced mRNA levels) a 30-fold TPA induced mRNA increase (Figure 1B: SA21/TIMP-1). The sequence of the cDNA SA11 corresponding to a 0.8 kb mRNA was found to be identical to an EST in the database of unknown function. This 0.8 kb mRNA demonstrated (in comparison to hemin induced levels) a 1.5 fold TPA induced increase.

### 3.2. Time course of TPA induction of K562 cells on expression of TIG-1 mRNA:

To determine whether there was a change in the steady state level of *TIG-1* mRNA with shorter periods of TPA induction, total cellular RNA isolated from K562 cells was analyzed at 1 hr, 4 hrs, 12 hrs, 24 hrs, 48 hrs and 72 hrs after induction with TPA. Northern blots were probed with a *TIG-1* cDNA fragment. The steady state

level of *TIG-1* mRNA was found to vary, with a 3.5-4 fold induction at 12 hrs in TPA induced cells in comparison to uninduced cells. The expression then declined to a 2-2.5-fold induction over 72 hrs. An increase in the steady state level of *TIG-1* mRNA occurs between 4-12 hrs following TPA induction (Figure 2A and 2B). Whether there is a cell cycle stage dependency on the level of *TIG-1* mRNA expression remains to be determined. The K562 cells were not synchronized prior to induction with TPA.

10

*Expression of TIG-1 mRNA in various tissues*

Northern blot analyses was used to determine whether the mRNA encoding *TIG-1* exhibited any tissue restricted pattern of expression. Screening was used to detect the presence of human *TIG-1* mRNA in hematopoietic and non-hematopoietic tissues. Using the 193 bp *TIG-1* cDNA fragment as a probe, a 3.7 kb transcript was detected in all tissues examined, as shown in Figure 3A. The wide range of tissues expressing *TIG-1* mRNA suggests that the protein encoded by this mRNA serves a constitutive function in the cell. However, the level of *TIG-1* mRNA expression varied; with the highest amounts of expression in the placenta and peripheral blood leukocytes, and lowest in the pancreas and the kidney (Figure 3B). In comparison to the kidney, there was 250 times more expression of *TIG-1* mRNA in the placenta and fifty times more expression in bone marrow derived tissues (Figure 3B). Whether there is the same wide variability in the level of *TIG-1* protein expression between tissues remains to be determined.

*Determination of the full length cDNA sequence of TIG-1*

Full length K562 cell and placental cDNA libraries were screened in order to obtain approximately 3.4 kb of nucleotide sequence. This contig contained 163 bp of 5' untranslated region, a 1,737 bp open reading frame  
5 encoding for 579 amino acids, and 1,499 bp of 3' untranslated region.

The proximal ATG codon begins at 164 bp from the start base pair of the 3.4 kb contig, and is preceded by the Kozak sequence GCCATGA in an adequate context,  
10 predicting a continuous open reading frame (ORF) of 1737 bp as shown in Figure 4. A putative polyadenylation signal AATAAA is located 16 bp upstream of the poly A tail and is underlined. The complete predicted ORF of the *TIG-1* cDNA sequence encodes a protein of 579 amino  
15 acids with a predicted molecular weight of 63 kDa. The deduced protein has an isoelectric pH of 10.0 and is rich in glutamine (28%), serine (10%) and proline residues (17%). The deduced amino acid sequence shows seven  
20 repeats of 6-16 glutamine residues, as shown in Figure 5. A putative bipartite nuclear localization signal was identified at amino acid residues 515-525 (Figure 5.).

While the full length sequence of *TIG-1* cDNA was being determined, a partial cDNA fragment CTG7a was found  
25 in the database which was identical to the *TIG-1* cDNA corresponding to amino acids 160-525 of the deduced open reading frame. This partial cDNA fragment was obtained by Ross et. al. 1997 following a screen of a brain cDNA library with a CAG repeat oligonucleotide in hopes of  
30 identifying novel genes involved in neurological disorders that are subject to trinucleotide expansion repeat mutations. Expression of *TIG-1* mRNA in brain tissue is demonstrated in Figure 3A.



*In vitro transcription and translation of TIG-1 cDNA*

To determine whether the *TIG-1* cDNA clone could produce a mRNA transcript that yielded a polypeptide of the predicted 63 kDa size, *in vitro* transcription and translation of *TIG-1* cDNA was carried out. Linearized *TIG-1* cDNA was transcribed by T3 RNA polymerase or T7 RNA polymerase, to generate antisense or sense transcripts respectively. These transcripts were then translated *in vitro* by a rabbit reticulocyte lysate system. The <sup>35</sup>S methionine labeled translated products were fractionated on a 10% SDS-PAGE gel. Translation of the sense transcripts produced a polypeptide of approximately 97 kDa and smaller polypeptides as seen in Figure 6A, lane 4. It is possible that the smaller polypeptides arise from incomplete translational products or are due to proteolytic degradation. Antisense transcripts failed to produce any polypeptide (Figure 6A, lane 5). No polypeptide bands were detected following *in vitro* translation of mRNA transcripts produced from a DNA template of the Bluescript plasmid ( Figure 6A, lane 6) or with a *TIG-1* cDNA template (Figure 6A, lane 1). The control, p45NF-E2 sense transcripts, yielded a 45 kDa polypeptide (Figure 6A, lane 2) while no translation products were detected with p45 NF-E2 antisense transcripts (Figure 6A, lane 3).

*Detection of TIG-1 protein in K562 cells:*

To detect *TIG-1* protein in K562 cells, an immuno-affinity purified rabbit polyclonal antibody was used to the carboxy terminal portion of bacterially expressed *TIG-1* protein in Western Blot analyses of cytosolic and nuclear extracts obtained from K562 cells. The affinity purified post-immune serum detected a

protein of 97 kDa in the nuclear and cytoplasmic extracts of uninduced K562 cells, as shown in Figure 6B, lanes 5 and lanes 6. No protein bands were detected with pre-immune serum (Figure 6B, lanes 1, 2, 3, and 4).

- 5 Similarly, the 97 kDa TIG-1 protein was detected in TPA-induced K562 cells with affinity-purified post immune serum (data not shown).

Western blot analyses of K562 protein lysates using affinity-purified antiserum preincubated with either  
10 bacterially expressed TIG-1 carboxy terminal peptide or bovine serum albumin confirmed the specificity of the antiserum. The 97 kDa TIG-1 protein band could not be detected by the affinity purified TIG-1 antiserum when it was preincubated with the TIG-1 carboxy terminal peptide;  
15 while the 97 kDa polypeptide band was observed when bovine serum albumin was preincubated with the affinity purified TIG-1 antiserum (data not shown).

Though the deduced open reading frame of the *TIG-1* cDNA predicts a polypeptide of MW 63 kDa both the *in*  
20 *vitro* transcription and translation data shown in Figure 6A, and the Western Blot analysis shown in Figure 6B demonstrate a protein product of MW 97 kDa. It is possible that the difference between the predicted and the observed MW of the TIG-1 protein is due to anomalous  
25 migration of a glutamine rich protein as observed by Cox et. al. 1996, or due to the high positive charge of the TIG-1 protein with a predicted pI of 10.0.

*CAT assay for detection of possible transcription  
30 regulatory activity of GAL4-TIG-1 fusion protein*

Though the deduced ORF of TIG-1 showed a putative consensus nuclear localization signal, no putative DNA binding domain was found. To determine whether the TIG-1

protein had any possible effect on activation of transcription cotransfection assays using a cat reporter vector and an expression vector encoding for the full length TIG-1 protein fused 3 minutes to DNA binding domain of GAL4 protein (GAL4-TIG-1). The cat reporter construct contained five GAL4 binding sites 5' to the minimal promoter for the adenovirus *E1B* gene, pGAL4<sub>5</sub>E1BCAT (Figure 7A). Expression of the GAL4-TIG-1 fusion plasmid would result in the tethering of TIG-1 protein to DNA via the DNA binding domain of GAL 4.

The *GAL4-TIG-1* expression plasmid cotransfected in the molar concentration of 1:5 to the CAT reporter plasmid showed a 3-fold increase in CAT expression in uninduced K562 cells, when compared to the CAT activity of the Gal4 DNA binding domain taken as 1 (data not shown). This suggests that TIG-1 protein could behave as a transcriptional activator. It was asked whether the glutamine rich domain of TIG-1 protein was sufficient for transcriptional activation (Gerber et. al. 1994). The construct *GAL4-GLU* which encodes for the glutamine rich portion of TIG-1 did not show any transactivation activity (Figure 7B and 7C, *GAL4-Glu*). It was verified that this fusion construct was indeed expressed by performing electrophoretic mobility shift analysis of nuclear extracts obtained from K562 cells transfected with this construct and a radiolabelled oligonucleotide probe containing the GAL4 operator sequence. The GAL4 operator probe was shifted by a GAL4:GLU fusion protein indicating that this fusion protein was made (data not shown). The reporter vector pGAL4<sub>5</sub>E1BCAT transfected in the absence of *GAL4-TIG-1* or with the expression vector for the GAL 4 DNA binding domain did not produce any cat gene expression (Figure 7B and 7C: neg and GAL4).

Strikingly, TPA induced K562 cells cotransfected with the CAT reporter construct and the GAL4-TIG-1 expression vector increased cat gene expression by 11-14 fold as compared to uninduced cells, after normalization for transfection efficiency and the non-specific effects of TPA, as shown in Figure 7. panel C: GAL4-TIG-1 open bar. No such effect of TPA induction on K562 cells was found with any of the other plasmids cotransfected in this experiment (Figure 7C except for GAL4-TIG-1). This result suggests that TPA induction of protein kinase C activity may result in regulation of the transcriptional activity of the TIG-1 protein. A putative protein kinase C phosphorylation site is found at amino acids 428-430 of the deduced open reading frame as depicted in Figure 4.

The results of the cotransfection assays suggest that the TIG-1 protein could function as a transcriptional activating factor, when tethered to DNA. Since a specific DNA binding domain could not be deduced, it is hypothesized that the TIG-1 protein may function as a transcriptional coactivator factor especially following induction of protein kinase C activity. Glutamine rich proteins have been observed as coactivators of transcriptions such as TAF110 (Hoey et. al.). Indeed, there are two classes of glutamine rich transcriptional regulatory proteins; one class including proteins such as TAF110 and TIG-1 contain clusters of glutamine residues, the other class including such proteins as nuclear respiratory factors 1 and 2 and Sp1 have glutamine containing clusters of hydrophobic residues (Gugneja et. al. 1996, Gill et. al. 1994). Glutamine rich proteins can form stable oligomers with other glutamine rich proteins as the polyglutamine tracts

Tjian's group identified a co-activator complex (ARC) that mediates chromatin-directed transcriptional

Using the affinity purified rabbit antiserum to the  
15 carboxy terminal 185 amino acids of TIG-1 protein  
reported in this paper, Naar et. al. was able to  
immunoprecipitate the ARC complex and demonstrate that it  
was identical to the ARC complex that was purified using  
(affinity chromatography for) the transcriptional  
20 activation domain of VP-16.

These transcriptional activator complexes appear to  
30 be found in all cells; furthermore there appears to be  
equal stoichiometry of the proteins within each complex.  
Though *TIG-1* mRNA appears to be ubiquitously expressed,  
data demonstrates that at least at the level of mRNA

expression there is substantial variation in the amount of *TIG-1* mRNA with highest levels of expression in placenta and bone marrow derived tissues. If indeed there is concomitant variation in expression of *TIG-1* protein, it is possible that *TIG-1* protein may participate in multiple complexes or in particular tissues to function in a tissue specific manner.

Whether *TIG-1* protein directly interacts with transcriptional activators or serves some other function within the transcriptional activator complex remains to be determined. Which members of ARC interact directly interact with *TIG-1* protein may be answered by expression of the *TIG-1* protein in a yeast two hybrid system.

Further studies to determine whether the activity of the *TIG-1* protein is regulated by phosphorylation are warranted based on the transfection data. Whether the *TIG-1* protein serves a unique function in the cell, or can be replaced by other members of the ARC complex will require gene knockout experiments.

It is of interest that the *TIG-1* mRNA encoding for a protein found to be part of a chromatin altering complex was independently identified using a subtractive hybridization protocol aimed at identifying novel genes encoding for proteins involved in differentiation of the human leukemia cell line K562. Expression of *TIG-1* mRNA was noted in hemin induced K562 cells, however, its expression was increased 2.5 fold when the K562 cells were induced with TPA. Since the phenotype of the K562 cell switches from erythroid to megakaryocytoid it is possible that the *TIG-1* protein/ARC complex is required for remodeling of chromatin structure at megakaryocytic specific promoters and enhancement of gene expression by a tissue specific subset of transcription factors.

Indeed, induction of K562 cells with hemin results in a very specific alteration of chromatin structure especially within and around the K562 cell  $\beta$ -globin gene domain-including alteration at the locus control region (LCR) leading to expression of epsilon and gamma globin genes (Tuan et. al., 1985); induction of K562 cells with TPA does not result in alteration of the chromatin structure of the  $\beta$ -globin LCR, extinguishes erythroid specific gene expression, while activating expression of other sets of genes-presumably requiring alteration of the chromatin structure, possibly by ARC/DRIP at the promoter/enhancers of these genes. Our data suggest that there is variation in the level of *TIG-1* mRNA expression in tissues with highest levels of expression in the placenta and bone marrow derived tissues; whether *TIG-1* protein participates complexes other than ARC/DRIP remains to be determined.

Transfection of the expression vector encoding for the GAL4:*TIG-1* fusion protein into K562 cells did result in some change in K562 cell morphology. Whether the *TIG-1* protein plays a role in the TPA induced megakaryocytoid differentiation of K562 cells remains to be determined.

Although preferred embodiments have been depicted and described in detail herein, it will be apparent to those skilled in the relevant art that various modifications, additions, substitutions and the like can be made without departing from the spirit of the invention and these are therefore considered to be within the scope of the invention as defined in the claims which follow.

# REFERENCES

- Alitalo, R., et al., "Enhanced expression of transforming growth factor beta during megakaryoblast differentiation of K562 leukemia cell line," BLOOD, 71: 899-901 (1988);
- Alitalo, R., et al. "Increased erythroid potentiating activity/tissue inhibitor of metalloproteinase and jun/fos transcription factor complex characterize tumor promoter induced megakaryocytic differentiation of K562 leukemia cells," BLOOD, 75: 1974-1982 (1990);
- Andersson, L.C., et al. "Induction of erythroid differentiation in the human leukemia cell line K562," Nature, 278: 364-365 (1990);
- Bayer, E.A., et al., Meth Enzym 62:308 (1979);
- Campbell, A.M., Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1984);
- Capecchi, M., Cell 22:479-488 (1980);
- Charnay, P. et al. "Transcriptional regulation of globin gene expression in the human erythroid cell line K562," Science 220: 1281-1283 (1983);
- Cox, G.W., et al. "Molecular cloning and characterization of a novel macrophage gene that encodes a nuclear protein comprising polyglutamine repeats and interspersing histidines" J. Biol. Chem., 271: 25515-25523 (1996);



Dean, A., et al. "Inducible transcription of five globin genes in K562 leukemia cells," Proc Natl Acad Sci USA, 80: 5515-5519 (1983);

5 Dignam, J., et al. "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei," Nuc. Acid. Res., 11: 1475-1489 (1983);

10 Engval, E., et al., Immunol 109:129 (1972);

Goding, J.W., J Immunol Meth 13:215 (1976);

Gerber H.P., et al. "Transcriptional activation modulated  
15 by homopolymeric glutamine and proline stretches," Science, 263: 808-811 (1994);

Gill, et al., "A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAF<sub>II</sub>110 component  
20 of the *Drosophila* TFIID complex and mediates transcriptional activation," Proc. Natl. Acad. Sci. USA, 91: 192-196 (1994);

Gugneja, S., et al., "Nuclear respiratory factors 1 and 2  
25 utilize similar Glutamine-containing clusters of hydrophobic residues to activate transcription," Mol. Cell. Biol., 16: 5708-5716 (1996);

Hoey, T., et al., "Molecular cloning and functional  
30 analysis of *Drosophila* TAF<sub>II</sub>10 reveal properties of coactivators," Cell, 72: 247-260 (1991);

5

10

15

20

25

Naar, A.M., et al., "Composite co-activator ARC mediates chromatin-directed transcriptional activation," Nature, 398: 828-83 (1999);

30

Pearson and Lipman, Proc Natl Acad Sci USA 85:2444  
(1988);

Perutz, M.F., "Glutamine repeats and inherited neurodegenerative diseases: molecular aspects," Current Opinion in Structural Biology, 6: 848-858 (1996);

- 5 Rachez, R., et al., "Ligand-dependent transcription  
activation by nuclear receptors requires the DRIP  
complex," Nature, 398: 824-828 (1999);  
Ross, L.M., et al., "cDNAs with long CAG trinucleotide  
repeats from human brain," Human Genetics, 100: 114-122  
10 (1997);

Sambrook et al., Molecular Cloning: A Laboratory Manual,  
2d Edition, Cold Spring Harbor Laboratory Press, Cold  
Spring Harbor, New York (1989);

- 15 Shigekawa, K. and Dower, W.J., BioTechniques 6:742-751  
(1988);

Siebert, P.D., et al., "Positive and Negative Regulation  
20 of Specific Gene Expression by the Tumor Promoter TPA in  
Human Erythroleukemic K562 Cells," Prog. Clin. Bio. Res.,  
191: 233-248 (1985);

Smith and Waterman, Adv Appl Math 2:482 (1981);

- 25 Sternberger, L.A., et al., J Histochem Cytochem 18:315  
(1970).

St. Groth, et al., J Immunol Methods 35:1-21 (1980).

- 30 Stott, K., et al., "Incorporation of Glutamine Repeats  
Makes Proteins Oligomerize: Implications For

Neurodegenerative Diseases," Proc. Natl. Acad. Sci. USA,  
92: 6509-6513 (1995);

Tuan D. et al., "β-like-gene Domain in Human Erythroid  
5 Cells," Proc. Natl. Acad. Sci. USA, 82: 6384-6388 (1985);

Wang, Z., et al., "A gene expression screen," Proc. Natl. Acad. Sci. USA, 88: 11505-11509 (1991).